DOCUMENT RESUME

ED 091 213 SE 017 783

AUTHOR Seymour, Lowell A.; And Others
TITLE The Measurement of Program Implementation and Students' Cognitive, Affective, and Social Performance in a Field Test of the Inquiry Role Approach (1972-73). III. Students' Cognitive, Affective and Social Skills Performance.
PUB DATE Apr 74
NOTE 28p.; Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (47th, Chicago, Illinois, April 1974). For related documents, see SE 017 781 and 782

EDRS PRICE MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS Affective Behavior; *Biology; *Cognitive Measurement; *Educational Research; *Instruction; Science Education; *Secondary School Science; Social Behavior; Teaching Methods
IDENTIFIERS *Inquiry Role Approach; Research Reports

ABSTRACT

This report is one of three concerning the 1972-73 field test of the Inquiry Role Approach (IRA) to biology teaching developed by the staff of the Mid-Continent Regional Educational Laboratory (McREL), Kansas City, Missouri. This paper contains a report of the students' cognitive, affective, and social skills performance. The 1,300 students participating in the study were measured by using the Comprehensive Final Examination; Exploration in Biology-Topic 1, Bird Populations; Biology Student Behavior Inventory; Social Skills Checklists; and Attitude Checklists. In addition, students completed two of the eight instruments (verbal reasoning, numerical ability) from the Differential Aptitude Test (DAT) battery, to provide a measure of general learning ability. The control group was found to be superior to verbal and numerical ability as measured by the DAT. Nevertheless, the IRA student groups had significantly superior posttest scores in cognitive inquiry and affective qualities of inquiry. The control group demonstrated significantly higher development in the area of biology content knowledge (BSCS Yellow Version) than did the IRA students. Results appeared to indicate that the IRA program was an effective teaching approach for developing cognitive inquiry skills and affective qualities of inquiry, both of which are considered important goals of science teaching. (Authors/PEB)

ED 091213

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE MEASUREMENT OF PROGRAM IMPLEMENTATION AND STUDENTS' COGNITIVE, AFFECTIVE, AND SOCIAL PERFORMANCE IN A FIELD TEST OF THE INQUIRY ROLE APPROACH (1972-73)

III. STUDENTS' COGNITIVE, AFFECTIVE AND SOCIAL SKILLS PERFORMANCE

by

Lowell A. Seymour
Richard M. Bingman
Paul G. Koutnik
Lawrence F. Padberg
Larry L. Havlicek
A. Thel Kocher
Kenneth A. Burton

A paper presented to the annual meeting of the National Association for Research in Science Teaching, Chicago, 1974.

McREL projects described in this report were supported in development by funds from the U. S. Office of Education and National Institute of Education, Department of Health, Education and Welfare, under contracts OEC-3-7-062876-3076 and NE-C-00-3-0068. The opinions expressed in this paper do not necessarily reflect the position or policy of the U. S. Office of Education or National Institute of Education and no official endorsement by these agencies should be inferred.

SE 017 783

Introduction --

The Inquiry Role Approach (IRA) is a method of teaching secondary biology which includes teacher training materials, teacher instructions for class use and student materials. Although the goals of IRA include the learning of biology content--factual information, concepts, and principles of biology--the goals emphasize inquiry skill development, social interaction skill, and attitude development necessary for good inquiry. The IRA method is based on the premise that biology content understanding, inquiry skills, social skills, and attitudes are interdependent and can be achieved best in a program that integrates them. The beginning point and developing rationale for this "four-pronged" approach have been reported previously (Seymour, et al., 1970; Bingman and Koutnik, 1970; and Koutnik, 1970).

Problems Studied --

The 1972-73 field test was undertaken to resolve four problems: Can the adequacy of IRA implementation be described in terms of teacher practices? Do students in classes in which IRA is implemented demonstrate the knowledge and skills which the program materials are designed to develop? Does student performance in IRA classes compare favorably with student performance in non-IRA classes? What recommendations for revision of program materials would be indicated by the field test? The first of these general problems was addressed in the first of the three papers in this paper set (Seymour, et al., 1974a). The sub-problem and hypotheses studied in the 1972-73 IRA field test, which relate to the remaining three general problems above, will be discussed in this paper. These sub-problems and hypotheses are:

SUB-PROBLEM 1: Have IRA students, in classes where the program was at least adequately implemented, shown significant increases from pre- to posttesting in biology content knowledge, cognitive inquiry skills, and affective qualities of inquiry [as measured by the Comprehensive Final Examination (Biological Sciences Curriculum Study, 1965), Explorations in Biology-Topic 1 (Koos, et al., 1972), and Biology Student Behavior Inventory (BSBI)(Steiner, 1970)]?

HYPOTHESIS 1: There is no significant gain from pre- to posttesting in biology content knowledge, cognitive inquiry skills, and affective qualities of inquiry for IRA students in classes where the program was at least adequately implemented.

SUB-PROBLEM 2: Have IRA students, in classes where the program was at least adequately implemented, met minimum performance levels for demonstration of social skills and affective qualities of inquiry at interim and posttesting [as measured by the Social Skills Checklists, IRA student forms 121-4 and 214-4, and the Attitude Checklists, IRA student forms 121-5 and 214-5 (Bingman, et al., 1972)]?

HYPOTHESIS 2: The mean scores of students in classes where IRA has been at least adequately implemented will not meet the criterion levels on the social skill and attitude checklists administered at interim and posttesting.

SUB-PROBLEM 3: Are there significant differences in IRA student outcomes in biology content knowledge, cognitive inquiry skills, and affective qualities of inquiry (measured by CFE, EIB-1, and BSBI, respectively) between students in the following groups: Students with verbal and numerical aptitude at the 75th percentile or above, from the 50th to the 74th percentile, from the 25th to the

49th percentile, and at the 24th percentile or below [percentiles based on Differential Aptitude Test-Verbal and Numerical scores (Bennett, et al., 1959)]?

HYPOTHESIS 3: There is no significant difference in IRA student outcomes in biology content knowledge, cognitive inquiry skills, and affective qualities of inquiry for students with different verbal and numerical aptitudes.

SUB-PROBLEM 4: Are there significant differences in student outcomes in biology content knowledge, cognitive inquiry skills, and affective qualities of inquiry (measured by CFE, EIB-1, and BSBI, respectively) between students in Inquiry Role Approach classes and students in non-Inquiry Role Approach classes?

HYPOTHESIS 4: There is no significant difference in student outcomes in biology content knowledge, cognitive inquiry skills, and affective qualities of inquiry among students grouped by: classes of IRA teachers using the BSCS Yellow Version text (Biological sciences Curriculum Study, 1968a), classes of experienced IRA teachers (that is, having previous experiences using IRA) using the BSCS Blue Version text (Biological Sciences Curriculum Study, 1968b), and classes of non-IRA teachers using the BSCS Yellow Version text.

SUB-PROBLEM 5: What revisions in the program materials are indicated by the results of testing, student feedback, and teacher feedback?

Choosing Participants --

During spring 1972, a letter seeking participants for the 1972-73 field test was sent to secondary biology teachers, school administrators, and other educators--college and university personnel, state boards of education personnel, etc. The field test would involve not only classroom teachers, but also trainers of teachers--department chairmen or curriculum supervisors--and, possibly, individuals such as university personnel to train the teacher trainers.

Accompanying the letter was a brief description of IRA and a questionnaire which sought such identifying information as whether or not the person was interested in participating in the field test, in what capacity, and if he could suggest additional persons to contact.

The initial mailing was sent March 22, 1972 to 47 persons in 16 states. Most individuals were in the McREL region (31 in Kansas, Missouri and Nebraska) and some had had previous involvement with the IRA program.

Lists of secondary biology teachers using the BSCS Yellow Version textbook in Missouri, Kansas and Nebraska were requested from the respective state departments of education. Partial lists were received and letters were sent to selected teachers (77 in Missouri, 10 in Kansas, 16 in Nebraska) during the month of April. It was found that the lists received were not current. Responses from these mailings were poor, apparently due to the dated information received from the state departments of education. About 10 additional teachers were contacted in various areas as a result of referrals returned to McREL by persons contacted in the initial mailings. Selection of participants from the questionnaire respondents was guided by the following criteria.

Guidelines for Selection of Field Test Participants --

1. A Distribution of Test Sites and a Variety of Trainers.

2. A Variety of Test Site Socio-economic Settings.

3. Heterogeneity of Student Abilities.

4. Adequate Sample Size: Krejcie and Morgan (1970) have developed a table based on a formula published by the National Educational Association (1960) for determining sample size in research activities. This table shows that the size of the teacher sample in the field test would not allow for generalization to a large teacher population. For example, a maximum of 40 teachers might be included; results with a sample of 40 can only be generalized to a population of 45. Therefore, our teacher sample size was determined by other factors--program staff and funding capabilities--rather than generalizability considerations.

   On the other hand, Krjecie and Morgan note: "As the population increases the sample size increases at a diminishing rate and remains relatively constant at slightly more than 380 cases." A selection of entries from Table 1 easily demonstrates this:

TABLE 1: Population Size Related to Sample Size

N (POPULATION SIZE)	S (SAMPLE SIZE)
1000	278
2000	322
5000	357
10000	370
20000	377
30000	379
40000	380
50000	381
75000	382
1000000	384

Therefore, to have the freedom of generalizing to almost any size population of similarly characterized secondary biology students, the student sample in the field test should be no less than 400. This figure was exceeded by 1,000 students.

As of July 31, 1972, the beginning of the IRA workshop at McREL, the field test participants included: (1) 4 teachers who would also train other teachers with 11 teachers using IRA but not responsible for training others; (2) 4 teachers without a trainer; (3) approximately 1,750 students in 65 class sections; and (4) 10 schools in 6 states.

In addition to these participants, eight teachers not using IRA materials, were asked to administer to their classes the battery of evaluation instruments used in the IRA classes. These teachers and their classes were the non-randomly assigned control group; approximately 465 students were included. These students

were similar to the test group IRA students in terms of heterogeneous grouping and other factors previously stated. The teachers were also similar to the test group teachers in terms of the textbook they used, experience in teaching, and general teaching approach. The primary difference was the lack of IRA materials and training for the control teachers. Pre- and posttests were administered in classes of four of the teachers; only posttests were administered in classes of the remaining four teachers.

Participating Teachers: All but 1 of the 15 field test participants had previous teaching experience and previous experience using the BSCS Yellow Version textbook. The teaching experience of the participants is summarized in Table 2.

TABLE 2: Years Teaching Experience of Field Test Participants

YEARS EXPERIENCE	NO. OF TEACHERS	TEACHER #
0 - 2	1	11
3 - 5	5	10, 12, 13, 31 & 22
6 - 9	2	02 & 01
10 - 15	3	04, 30 & 14
16 or more	4	40, 20, 21 & 03

Note that the one inexperienced teacher worked in a team teaching setting with four other experienced teachers.

These teachers have also been categorized according to the type of IRA training they received. (See the first paper of this paper set.)

Participating Students: IRA is designed for students with abilities and achievement in the 30th to 99th percentile range as measured by the Differential Aptitude Test-Verbal and Numerical. Inclusion of students falling below the 30th percentile should not affect the success of the program, neither overall or for those students below the 30th percentile, as long as the student groups are heterogeneous and the percentage of students below the 30th percentile remains low.

Mean percentile for students in the 1972-73 field test, according to DAT Verbal and Numerical scores, are given in tables 3 and 4.

TABLE 3: Mean Percentile for Scores on DAT-Verbal

SITE	N	MEAN RAW SCORE	MEAN PERCENTILE*
A	97	34.27	75
B	508	28.22	57
C	203	29.89	63
D	203	31.77	68
E	141	24.03	43
F	131	27.24	55
G	51	28.47	58
H	19	31.79	68
Total	1353	29.01	60

*10th grade, first semester norms applied.

TABLE 4: Mean Percentile for Scores on DAT-Numerical

SITE	N	MEAN RAW SCORE	MEAN PERCENTILE*
A	94	26.56	63
B	456	18.11	30
C	206	23.13	48
D	195	24.48	53
E	124	17.73	27
F	129	21.57	42
G	51	22.51	45
H	18	22.89	47
Total	1273	21.08	40

*10th grade, first semester norms applied.

The DAT-V mean for the entire IRA sample was 29.01, as reported in Table 3, and the percentile rank for this mean was 60. The median was 29.00 and the mode was 28.00. Thus these scores were probably normally distributed. The minimum score was 7 and the maximum was 48 on this 50-item test. The standard deviation was 9.94. Thus 84.38 percent of the IRA students had verbal scores at or above the 30th percentile on the DAT-Verbal test.

The mean, median, mode, minimum, maximum, and standard deviations were 21.12, 21.00, 18.00. 1.9. 40.0, and 8.18, respectively, for all IRA students on DAT-N. Thus 64.4 percent of the IRA students had numeric scores at or above the 30th percentile on the DAT-Numerical test.

All students were in their first year biology classes using the BSCS Yellow Version text. Students at Site E were ninth graders; at Site A, students were primarily 11th graders; at all other sites, students were all, or primarily, tenth graders. A large percent of students at Sites B, E, F, and G were below the 30 percentile range. This was higher than preferred.

Control Group Participants: All teachers in the control groups were experienced teachers. Classes included were first year biology using BSCS Yellow Version texts composed of all or primarily tenth grade students. Four of the control teachers tested at the beginning and end of the school year; four others tested only at the end of the year.

A description of the student populations at the control sites according to DAT scores is given in Tables 5 and 6.

TABLE 5: Control Students Percentile Group Distribution According to DAT-Verbal Scores

SITE	STUDENT N	MEAN RAW SCORE	MEAN PERCENTILE*
A	145	35.41	77
C	55	29.78	65
E	51	29.53	63
H	66	30.98	65
I	148	33.10	70

*10th grade, first semester norms applied.

TABLE 6: Control Students Percentile Group Distribution According to DAT-Numerical Scores

SITE	STUDENT N	MEAN RAW SCORE	MEAN PERCENTILE*
A	145	29.21	75
C	55	24.98	57
E	51	24.55	55
H	66	24.47	55
I	148	25.95	62

*10th grade, first semester norms applied.

Experienced IRA Teachers: Four teachers in the Kansas City area have participated for five years (1968-69 through 1972-73) in the testing and development of the Inquiry Role Approach program. They were experienced with prototype IRA materials keyed to the BSCS Blue Version text. During the 1972-73 school year, these teachers adapted the IRA field test materials, keyed to the BSCS Yellow Version text, to the Blue Version text. While the emphasis of the field test focused on the results in classes of teachers using the Yellow Version text, evaluation instruments were administered to student samples of each of these experienced teachers when possible.

A description of the students populations of these experienced IRA teachers according to DAT scores is given in Tables 7 and 8.

TABLE 7: Experienced IRA Teachers' Students Percentile on DAT-Verbal Scores

TEACHER	STUDENT N	MEAN RAW SCORE	MEAN PERCENTILE*
61	17	30.1	63
62	28	32.5	70
63	72	31.8	69
64			

TABLE 8: Experienced IRA Teachers' Students Percentile on DAT-Numerical Scores

TEACHER	STUDENT N	MEAN RAW SCORE	MEAN PERCENTILE*
61	17	23.4	49
62	28	26.7	64
63	72	24.2	53
64			

Students were in first year biology. Students of Teacher 63 were all in ninth grade. Students of the other teachers were all or primarily tenth graders.

Instruments --

The Comprehensive Final Examination, Exploration in Biology-Topic 1, Bird Populations, and Biology Student Behavior Inventory have been described in the first paper of this paper set. The Social Skills Checklists and Attitude Checklists have been described in the second paper of this paper set (Seymour, et al., 1974b).

Differential Aptitude Test: The DAT (Bennett, et al., 1966) is a battery of instruments designed to measure student aptitude in eight areas. Two of the eight instruments -- Verbal Reasoning and Numerical Ability -- are often used together as a measure of general learning ability (DAT manual, p. 1-7). Only

these two instruments of the DAT battery were used in the field test. These were administered in the fall of the year to establish a base for comparison made between groups in the field test and for comparing the field test group as a total with outside populations.

Validity: A large number of studies have been performed relating course grades for various subjects to DAT scores. It was adequate for our purposes to note that of the coefficients of correlation computed for science grades compared to the nine DAT scores (8 instruments and the Verbal Reasoning + Numerical Reasoning composite score), the highest coefficients were found for Verbal Reasoning (.45), Numerical Ability (.44) and the VR+NA composite (.52).

Validation by a 3-1/2 year longitudinal study was also performed. This study indicated that DAT scores remain predictive of student performance over a long range. For example, DAT VR and NA scores from students 8th grade (mid-year) correlated well with general science grades achieved at end of 8th grade (VR - science grades, r = .64; NA - science grades, r = .59); these 8th grade DAT scores still correlated well with science (physics) grades achieved at end of 11th grade (VR - physics grades, r = .59; NA - physics grades, r = .60).

A most important means of validating the DAT was in appraising its predictive ability of student results on achievement tests. Some examples of the coefficients of correlation found between DAT-VR, DAT-NA and DAT-VR+NA scores and various achievement tests are given in the following table:

TABLE 9: Coefficients of Correlation Between DAT-VR, DAT-NA, and DAT-VR+NA scores and various achievement tests

TEST	COEFFICIENTS OF CORRELATION							
	BOYS				GIRLS			
	N	VR	NA	VR+NA	N	VR	NA	VR+NA
Iowa Test of Basic Skills - Form 1 - Reading Comprehension	125	.62	.61	.69	117	.68	.61	.73
Arithmetic Total	125	.71	.69	.80	117	.53	.75	.76
Iowa Tests of Educational Development - Form Y4-FL Composite	93	.91	.85	.92	79	.89	.76	.89
Stanford Achievement Test - Form KM, Intermediate Level - Battary Median	74	.84	.84	.91	71	.82	.90	.92

In general, the DAT scores have shown high correlations with achievement tests measuring comparable skills and knowledge.

Reliability: Reliability was studied using the split half technique with the computed correlation coefficients corrected by the Spearman-Brown formula. The VR, NA, and VR+NA coefficients (given separately for form L and M, for boys and

girls, and for each grade 8 through 12) range from .83 to .96. The tenth grade values for Form L are: for boys, Verbal Reasoning, r = .93, Numerical Ability, r = .91, VR+NA, r = .95; for girls, Verbal Reasoning, r = .94, Numerical Ability, r = .91, VR+NA, r = .96.

The long term consistency of measurement by the DAT was studied by determining the correlation between 9th grade scores and 12th grade scores for the same set of students studied over the three year period. Verbal Reasoning coefficients of correlation were .87 for boys (N = 71) and .82 for girls (N = 90); Numerical Ability coefficients for these same groups were .75 for boys, .74 for girls. This study utilized DAT - form A.

Correlation to other tests: The DAT correlates well with most standard intelligence tests. Some examples of the coefficients of correlation found between DAT-VR, DAT-NA and DAT-VR+NA scores and various intelligence tests are given in the following table:

TABLE 10: Coefficients of Correlation Between DAT-VR, DAT-NA, and DAT-VR+NA Scores and Various Intelligence Tests

TEST	COEFFICIENTS OF CORRELATION							
	BOYS				GIRLS			
	N	VR	NA	VR+NA	N	VR	NA	VR+NA
Lorge-Thorndike intelligence tests (Form A, Level 4) - Taken in 11th grade,								
Verbal	58	.70	.60	.72	59	.85	.78	.86
Non-Verbal	58	.61	.57	.64	59	.72	.69	.74
School and College Ability Tests (Form 2A								
Verbal	71	.82	.57	.78	59	.83	.64	.80
Quantitative	71	.67	.83	.81	59	.77	.82	.85
Total	71	.85	.79	.90	59	.87	.77	.89

Data Analysis and Interpretation--

Data Processing: The general sequence of data processing was as follows:

1. Distribution of measuring instruments and instructions to field test participant teachers.
2. Administration of instruments by teachers.
3. Collection of data by McREL.
4. Scanning or key punching data onto cards.

5. Scoring of instruments.

6. Analysis of scores per various groups of subjects.

This basic sequence was repeated three times during the field test to obtain pretest data, interim data after Theme I and posttest data. A brief description of data collected and the approximate times these data were collected are indicated in Chart I.

The statistical processing of data collected during the field test was performed on computers located at the University of Missouri-Columbia (IBM 370/165) and at the University of Kansas (Honeywell 635). For information concerning the particular programs used for the different analyses performed, see Table 11. In a few instances, post hoc analyses were computed on desk calculators. All analyses were performed using the student as the sampling unit.

CHART I: Data Collection for IRA Field Test 1972-73

PRETESTING
September, 1972

Differential Aptitude Test - Verbal Reasoning and Numerical Ability
Comprehensive Final Examination - Form J
Exploration in Biology-Topic 1. Bird Population
Biology Student Behavior Inventory

INTERIM (END OF THEME I) TESTING
December, 1972 - January, 1973

Class Activities Questionnaire
Views and Preferences - Form C
Explorations in Biology - Topic 2. Food Preferences of Newly-Hatched Snakes
Social Skills Checklist (IRA student form 121-4)
Attitude Checklist (IRA student form 121-5)
Understanding Role Responsibilities (IRA student form 121-3)
A biology content test designed by the teacher

POSTTESTING
May-June, 1973

Comprehensive Final Examination - Form K
Explorations in Biology - Topic 1. Bird Populations
Biology Students Behavior Inventory
Class Activities Questionnaire
Views and Preferences - Form C
Social Skills Checklist (IRA student form 308-1, 3 teachers; IRA student form 214-4, 11 teachers)
Attitude Checklist (IRA student form 308-2, 3 teachers; IRA student form 214-5, 11 teachers)

TABLE 11: Listing of Computer Programs Used for Data Analyses

This PROGRAM was used to obtain this ANALYSIS to support this HYPOTHESIS.*		
DATSCOR		
BSBSCOR		
EIBSCOR	SCORED	
PARTPUN	OUTPUT	
SORT (U)		
CONDENS		
MISDATA	Analysis of Variance	1; Pre-sensitization
BMD04V	Analysis of Covariance and Newman-Keuls Post Hoc analysis*	3
SFA41D	Correlations	Correlations
TESTAT	ITEM Analysis	Reliability data for EIB & BSBI
ANOVAR1	Analysis of variance and Newman-Keuls A Posteriori analysis	4
VAPSCOR	Mean, Criterion level classification**	
SUMCTAB	Descriptive statistics	2

* Also used for study of student outcomes vs. degree of implementation reported in paper one of this paper set.
**Used for description of degrees of implementation reported in paper one.

SUB-PROBLEM 1: Have IRA students, in classes where the program was at least adequately implemented, shown significant increases from pre- to posttesting in biology content knowledge, cognitive inquiry skills, and affective qualities of inquiry (as measured by the Comprehensive Final Examination, Explorations in Biology-Topic 1, and Biology Student Behavior Inventory)?

HYPOTHESIS 1: There is no significant gain from pre- to posttesting in biology content knowledge, cognitive inquiry skills, and affective qualities of inquiry for IRA students in classes where the program was at least adequately implemented.

Data Analysis/Results: In order to determine whether or not there were any significant gains from pretest to posttest for any of the student outcome variables, an analysis of variance, non repeated measures, was computed for each variable. The results of these analyses are presented in Table 12. Note that this objective and hypothesis dealt only with students in classes where IRA was at least adequately implemented. Therefore, data from teacher 01 were not included in any of these analyses.

TABLE 12: Number of Students, Pretest and Posttest Means, F Ratios, and Probability Levels for Student Outcome Variables (Analysis of Variance, Non-Repeated Measures)

VARIABLE	PRETEST MEAN	N PRE	POSTTEST MEAN	N POST	F RATIO	P
EIB IA*	18.93	568	20.63	812	36.1	.0000
EIB IB**	35.86	573	40.90	786	113.6	.0000
BSBI A Curiosity	2.60	580	2.71	519	7.53	.006
BSBI B Openness	3.52	580	3.70	519	15.66	.0003
BSBI C Satisfaction	3.58	580	3.53	519	1.74	.18
BSBI D Responsibility	3.55	580	3.85	519	21.57	.0000
BSBI Total Score	13.25	580	13.79	519	17.08	.0002
CFE	17.56	589	19.64	777	40.39	.0000

* EIB-1A is a subscore of EIB 1 which includes EIB subscales I, III. and 12 items from subscale IV.

** EIB-1B is a subscore of EIB 1 which includes 12 additional items from subscale IV and subscales V and VI.

As can be noted from Table 12, seven of the eight F ratios were significant beyond the .01 level. All of the differences were in a positive direction. Thus these analyses indicate the null hypothesis can be rejected for all variables except BSBI subscale C (satisfaction).

Inquiry Role Approach students, in classes where at least adequate implementation had occurred, scored significantly (P = less than .01) higher at the end of the school than at the beginning for: cognitive inquiry skills as measured by EIB-1A and EIB-1B; affective qualities of inquiry as measured by the BSBI total score and subscale A (Curiosity), B (Openness) and D (Responsibility); and biology content knowledge as measured by the CFE.

The design utilized for testing the hypothesis was a quasi-experimental design. Campbell and Stanley (1963) have noted that this design may be appropriate in field situations where equivalent or comparable control groups cannot be added. It is further characterized as tending toward superiority in external validity or generalizability over "true" experimental designs. However the most important characteristic of this design for the purposes of this study was its ability to control for the effect of taking a pretest upon the scores of a posttest.

It should be noted that the design used here did not control for maturation--pre to post changes resulting from the passage of time rather than treatment. However, a modified Solomon Four-Group Design was used for Problem 4, and posttest only analyses were performed comparing experimental and control groups.

For pretesting, students were randomly distributed into two groups. Group 1 was pretested with the BSBI and CFE instruments; Group 2 was pretested with the EIB-Topic 1 instrument. All students were posttested with all three instruments.

Thus Group 2 students acted as a non-pretested control group for the BSBI and CFE instruments; Group 1 students acted as a non-pretested control group for the EIB-1 instrument. An analysis of variance was computed between those students who had the pretest for each variable and those students who did not have the pretest. The results of these analyses are presented in Table 13.

TABLE 13: Posttest Mean Scores and F Ratios for Comparison of Students With and Without Pretests

TEST	PRETESTED (GROUP 1) MEAN	N	NOT PRETESTED (GROUP 2) MEAN	N	F	P
EIB III	11.20	399	11.41	432	.67	.58
EIB IV	17.61	339	17.36	361	.95	.67
EIB V	23.96	398	23.24	426	3.40	.06
EIB VI	6.81	387	6.82	415	.02	.89
EIB Total	60.63	339	60.35	361	.11	.74
TEST	PRETESTED (GROUP 2) MEAN	N	NOT PRETESTED (GROUP 1) MEAN	N	F	P
BSBI A	2.70	214	2.70	378	.02	.89
BSBI B	3.64	214	3.63	378	.02	.89
BSBI C	3.50	214	3.57	378	1.38	.24
BSBI D	3.67	214	3.71	378	.17	.69
BSBI Total	13.51	214	13.60	378	.18	.67
CFE	18.74	406	19.20	393	1.02	.31

The results indicate that there were no significant differences between the two groups on any of the posttest scores. As can be noted in Table 13, the mean for the two groups were very close and in all cases except for the EIB 3 part score, the group which did not have the pretest scored slightly but not significantly higher than the group of students who had the EIB test as a pretest. For the BSBI scores, the mean were again very close and the group of students with the BSBI as a pretest scored slightly but not significantly higher on two of the part scores and the total score. As noted above, none of these differences were significant at the .05 level of significance. For the CFE, the group of students who had the CFE as a pretest scored about half a point higher than the group of students who did not have this test as a pretest, but again the difference was not significant at the .05 level.

SUB-PROBLEM 2: Have IRA students, in classes where the program was at least adequately implemented, met minimum acceptable performance levels for demonstration of social skills and affective qualities of inquiry at interim and posttesting (as measured by the Social Skills Checklists, IRA student forms 121-4 and 214-4, and the Attitude Checklists, IRA student forms 121-5 and 214-5)?

HYPOTHESIS 2: The mean scores of students in classes where IRA has been at least adequately implemented will not meet the criterion levels on the social skills and attitude checklists administered at interim and posttesting.

Analyses/Results: The social skills and attitude checklists have been developed to be specific measures for the sets of social skills and attitudes which the IRA program is designed to foster in students. These are unlike the non-IRA specific instruments discussed previously (Comprehensive Final Examination, Explorations in Biology, and Biology Student Behavior Inventory) and are used to measure student pre-to-post gains in biology content knowledge, cognitive inquiry skills, and selected inquiry attitudes. The social skills and attitude checklists are criterion referenced measures. It is felt that the use of these instruments would not be particularly appropriate as pre-to-post gain or IRA vs. Non-IRA measures.

The development of these instruments has been discussed in the second paper of this paper set (Seymour, et al., 1974b).

All teachers administered the Theme I social skills and attitude checklists (121-4 and 121-5). Table 14 presents the data from this testing.

TABLE 14: Social Skill Checklist and Attitude Checklist Data from End of Theme I

TEACHER NO.	Students' Mean Score Social Skill Checklist 121-4		Students' Mean Score Attitude Checklist 121-5	
	$\bar{X}$	N	$\bar{X}$	N
02	43.2	24	48.4	23
03	40.2	22	46.6	22
04	49.4	26	*	*
20	33.4	77	41.7	82
21	35.9	20	44.4	20
22	41.0	20	45.6	20
23	42.4	26	49.4	26
30	38.8	26	43.0	27
31	43.2	25	51.1	25
40	42.0	103	47.0	103
Criterion	28.0		33.0	

* No data submitted.

Teachers sent raw data (students' checklists) to the IRA staff. Random samples (20-25 papers per teacher) were taken from each teacher's data submitted for calculating these means. Note that teacher 40 had calculated mean scores for all 103 students completing the instruments. Note that in all cases students' mean score exceeded the criterion level.

Note that teacher 01, who did not adequately implement the program, is not included in this data. However this teacher's students did meet criterion on these instruments (121-4, $\bar{X}$ = 33.8; 121-5, $\bar{X}$ = 43.7).

Teachers varied in the number of IRA activities each completed. (See paper one in this paper set.) As a result, posttesting with social skill and attitude checklists was not uniform. Some teachers administered Theme II checklists

(214-4 and 214-5) during the second semester but before the posttesting. Some administered Theme II checklists as part of the posttesting. And at least one teacher administered the Theme III checklists (308-1 and 308-2) as part of the posttesting, having administered the Theme II checklists earlier in the semester.

Responsibility was given to the teachers to summarize social skill and attitude checklist data collected during the second semester, rather than having raw data submitted. In retrospect, data collection emphasis was placed on those measures used for pre-to-post gain and IRA vs. non-IRA comparison studies. Social skill and attitude checklist data was not properly reported by teachers. Eight teachers (02, 03, 10, 11, 12, 13, 14, and 21) submitted raw data (student papers) for Social Skills Checklist 214-4 and Attitude Checklist 214-5. A random sample of 50 was selected for each instrument; data from this sample is given in Table 15.

TABLE 15: Social Skill Checklist and Attitude Checklist Data from End of Theme II

	$\bar{X}$ Score	N	Criterion Score	S	T
Social Skills Checklist 214-4	71.8	50	53	13.3	10.1*
Attitude Checklist 214-5	58.3	50	43	11.4	9.5

* P .0005 one tailed

Thus the mean scores for a student sample representing nine of the 14 teachers who adequately implemented the program well exceeded the criterion levels. Among the 50 students in the social skills checklist sample, only 3 (6%) failed to meet criterion. In the attitude checklist sample, 4 (8%) failed to meet criterion. Two t-tests were calculated to determine if the mean scores were significantly larger than the criteria levels. Results (see Table 15) were such that the null hypothesis, hypothesis 2, can be rejected. IRA students did meet criterion levels on the social skill and attitude checklists.

SUB-PROBLEM 3: Are there significant differences in IRA student outcomes in biology content knowledge, cognitive inquiry skills, and affective qualities of inquiry (measured by CFE, EIB-1, and BSBI, respectively) between students in the following groups: Students with verbal and numerical aptitude at the 75th percentile or above, from the 50th to the 74th percentile, from the 25th to the 49th percentile, and at the 24th percentile or below (percentiles based on Differential Aptitude Test-Verbal and Numerical scores)?

HYPOTHESIS 3: There is no significant differences in IRA student outcomes in biology content knowledge, cognitive inquiry skills, and affective qualities of inquiry for students with different verbal and numerical aptitudes.

Data Analyses/Results: The analysis of covariance was used to determine whether or not there were any significant differences in student outcomes variables (EIB-subscale I not included) among the four subgroups based on both the DAT-Verbal and the DAT-Numerical scores. Pretest scores were held constant for

each variable analyzed. The results of these analyses are presented in Tables 16 and 17. The Newman-Keuls analysis was used to determine which pairwise differences were significant; these results are presented in Tables 18 and 19.

TABLE 16: Adjusted Means and F Ratios for Comparing Student Subgroups Based on Quartiles on DAT-Verbal Reasoning

VARIABLE	FIRST QUARTILE ADJUSTED MEAN	N	SECOND QUARTILE ADJUSTED MEAN	N	THIRD QUARTILE ADJUSTED MEAN	N	FOURTH QUARTILE ADJUSTED MEAN	N	F RATIO
EIB III	10.19	35	10.07	51	12.02	82	12.78	117	13.06**
EIB IV	15.69	27	16.83	33	17.66	66	19.00	90	10.54**
EIB V	18.89	38	20.26	42	23.51	93	27.27	109	40.25**
EIB VI	5.80	34	6.87	38	6.76	88	7.61	106	9.67**
EIB Total	53.59	27	54.45	33	60.79	66	68.17	90	34.38**
								82	
BSBI A	2.54	21	2.52	48	2.67	65	2.77	82	1.90
BSBI B	3.45	21	3.43	48	3.60	65	3.76	82	3.05*
BSBI C	3.38	21	3.37	48	3.49	65	3.65	82	2.98*
BSBI D	2.91	21	3.35	48	3.72	65	3.90	82	6.52*
BSBI Total	12.46	21	12.76	48	13.51	65	13.94	82	5.26*
CFE	16.11	25	18.41	45	17.27	72	19.89	98	3.68*

* Significant at .05 level
** Significant at .01 level

TABLE 17: Adjusted Means and F Ratios for Comparing Student Subgroups Based on Quartiles on DAT-Numberial Ability

VARIABLE	FIRST QUARTILE ADJUSTED MEAN	N	SECOND QUARTILE ADJUSTED MEAN	N	THIRD QUARTILE ADJUSTED MEAN	N	FOURTH QUARTILE ADJUSTED MEAN	N	F RATIO
EIB III	11.11	86	11.72	67	12.01	96	13.06	21	2.41
EIB IV	16.18	61	17.32	52	18.79	81	19.69	19	13.03**
EIB V	21.09	85	22.99	75	25.96	94	28.13	22	17.81**
EIB VI	6.63	76	6.82	72	7.29	41	7.96	20	4.00*
EIB Total	56.41	61	60.26	52	65.22	81	69.26	19	14.64**
BSBI A	2.44	52	2.66	68	2.74	78	2.85	17	3.36*
BSBI B	3.46	52	3.59	68	3.72	78	3.52	17	1.97
BSBI C	3.40	52	3.51	68	3.53	78	3.65	17	.94
BSBI D	3.21	52	3.63	68	3.85	78	3.80	17	4.51*
BSBI Total	12.58	52	13.41	68	13.81	78	13.65	17	4.45*
CFE	16.21	67	17.80	79	19.82	75	21.62	18	6.35*

* Significant at .05 level
** Significant at .01 level

TABLE 18: Newman-Keuls Post Hoc Analysis for DAT-Verbal Quartiles

QUARTILE PAIRINGS:	1ST / 2ND	1ST / 3RD	1ST / 4TH	2ND / 3RD	2ND / 4TH	3RD / 4TH
EIB 3		**	**	**	**	
EIB 4	**	*	*		*	
EIB 5	**	**	**	**	**	
EIB 6			**		**	**
EIB Total	**	**	**	**	**	
BSBI B		*				
BSBI C						
BSBI D		*	**		**	
BSBI Total		*	**		*	
CFE			*			

* Significant at .05 level
** Significant at .01 level

TABLE 19: Newman-Keuls Post Hoc Analysis for DAT-Numerical Quartiles

QUARTILE PAIRINGS:	1ST / 2ND	1ST / 3RD	1ST / 4TH	2ND / 3RD	2ND / 4TH	3RD / 4TH
EIB 4		*	*	**	*	
EIB 5		**	**	**	**	
EIB 6		*	**			
EIB Total		**	**	*	**	
BSBI D			*		*	
BSBI Total						
CFE		*	**			

* Significant at .05 level
** Significant at .01 level

As indicated in Tables 16 and 17, all but one of the F ratios for the total and four subscale scores on the EIB were significant; all but three of the F ratios for the total and four subscale scores on the BSBI were significant; and the F ratios for the CFE were significant. Tables 18 and 19 indicate which of the pairwise comparisons were significant. It should be noted that, although the F ratios were significant for BSBI-subscale C compared to DAT-Verbal and BSBI-total score compared to DAT-Numerical, the Newman-Keuls analysis did not result in any significant pairwise differences.

In order to further clarify the possible relationships between student outcome variables and DAT scores, correlation coefficients were computed between

each measure of student outcome and the DAT scores. Table 20 presents the results of this analyses.

TABLE 20: Correlations Between Posttest Student Outcome Variables and DAT-Verbal and DAT-Numerical Scores

	r* DAT-V	N	r* DAT-N	N
EIB 3	.417	742	.318	718
EIB 4	.450	636	.425	623
EIB 5	.550	735	..468	722
EIB 6	.361	716	.278	703
EIB Total	.610	636	.525	623
BSBI A	.242	522	.249	499
BSBI B	.484	522	.435	499
BSBI C	.299	522	.294	499
BSBI D	.462	522	.428	499
BSBI Total	.507	522	.479	499
CFE	.481	717	.482	687

* All Correlations are significant at 0.0

Interpretation: It is apparent from the Newman-Keuls test results shown in Tables 18 and 19 that student outcomes in cognitive inquiry as measured by the instrument EIB-Topic 1 were related to both DAT-Verbal and Numerical scores since there were a number of significant differences between the various quartile subgroups. The correlation coefficients for EIB-Total scores (the coefficients indicating significant positive linear relationships) also support this view.

Tables 18 and 19 also show that student outcomes for affective qualities measured by the BSBI were related to DAT-Verbal scores. Only two pairwise comparisons for BSBI - subscale D show significant differences; BSBI-total scores show no significant differences in pairwise comparisons. Therefore there does not appear to be a substantial relationship between BSBI and DAT-Numerical. The correlation coefficient (.479) would support this view. This is as expected since the BSBI instrument is designed to measure affective qualities.

In the comparison of CFE to DAT-Verbal, only one quartile pairing, 1st to 4th, showed a significant difference ($p = .05$). Two pairings showed significant differences when CFE and DAT-Numerical were compared (1st to 3rd, $p = .05$; 1st to 4th, $p = .01$). CFE and DAT scores therefore were apparently related, but not to the degree shown for EIB and DAT scores. This view is again supported by the correlation coefficients ($r = .481$, CFE-DAT-Verbal; $r = .482$, CFE-DAT-Numerical).

SUB-PROBLEM 4: Are there significant differences in student outcomes in biology content knowledge, cognitive inquiry skills, and affective qualities of inquiry (measured by CFE, EIB-1, and BSBI, respectively) between students in Inquiry Role Approach classes and students in non-Inquiry Role Approach classes?

HYPOTHESIS 4: There is no significant difference in student outcomes in biology content knowledge, cognitive inquiry skills, and affective qualities of inquiry among students grouped by: Classes of IRA teachers using the BSCS Yellow Version text, classes of experienced IRA teachers (that is, having previous experience using IRA) using the BSCS Blue Version text, and classes of non-IRA teachers using the BSCS Yellow Version text.

Data Analyses/Results: It is important to first identify which teachers' students were included for these analyses. As DAT scores became available it was readily noticed that the DAT mean scores for students in the three groups given above (IRA Yellow Version classes, IRA Blue Version classes, non-IRA Yellow Version classes) were not equal. Particularly, IRA Yellow Version classes were well below the other student groups. Since it would be inappropriate to simply eliminate selected students with low DAT scores from the analyses, a decision was made to delete groups of students with low DAT mean scores. Thus teacher 01's students (mean score DAT-Verbal = 24.03; mean score DAT-Numerical = 17.73) were deleted as a group. (It should also be noted that teacher 01 did not meet criteria for adequate IRA implementation, and therefore student outcomes would not be considered valid IRA results.) In addition, teacher group 10's students (mean score DAT-Verbal = 28.22; mean score DAT-Numerical = 18.11) were deleted as a group. (Teacher group 10 represented a unique team teaching implementation design with no matching control group on this variable.) These deletions raised the IRA Yellow Version students' mean DAT scores from 29.01 to 30.71 on the Verbal and from 21.12 to 23.87 on the Numerical. This was an increase from approximately the 60th to 65th percentile on the Verbal and from the 40th to the 50th percentile on the Numerical (using 10th grade, first semester norms). Therefore all analyses using IRA Yellow Version scores include data from students of all teachers except teacher 01 and teacher group 10.

Students' scores from all eight control teachers (non-IRA Yellow Version) were included in the EIB and CFE analyses. Three teachers did not administer the BSBI.

The control group (students of all eight teachers) had a mean DAT-Verbal score of 32.73 (70th percentile on 10th grade first semester norms) and a mean DAT-Numerical score of 26.40 (63rd percentile). These DAT mean scores were not significantly different for the students of the five teachers included in the BSBI analysis.

In order to determine if the primary experimental (IRA-Yellow) group student means for verbal and numerical ability were different from the respective means for the control group, a t-test was utilized. The results are shown in Table 21.

TABLE 21: Comparison of IRA and non-IRA Yellow Version Students' DAT-Verbal and Numerical Mean Scores

	DAT - ($\bar{X}$)		S.D.		N			
	IRA	N-IRA	IRA	N-IRA	IRA	N-IRA	t	P
Verbal	30.71	32.73	9.05	8.99	668	487	3.74	.01
Numeric	23.87	26.40	7.28	7.49	656	487	5.62	.01

Thus the control (non-IRA Yellow Version) group had significantly superior DAT-Verbal and Numerical ability over the experimental group (IRA-Yellow Version) used in the following analyses. However, percentile comparisons, as noted earlier, were improved by the deletion of teachers 01 and 10. Further depletion of the experimental group to raise mean DAT scores would not be greatly improved unless a large number of groups were deleted.

The experienced IRA Blue Version teachers reported a student DAT-Verbal mean score of 31.72 (68th percentile) and a student DAT-Numerical mean score of 24.68 (55th percentile). Teacher 64 did not report DAT scores but it was assumed his students are nearly the same since they are within the same district as students of teachers 61 and 62. Note that CFE and BSBI analyses included students from all four of these teachers. EIB analyses, however, included data from one teacher, 64; the others did not administer the EIB instrument.

In order to determine if there were any significant differences among three groups of teachers' students on any of the posttest scores, a one-way analysis of variance was applied to each of the student outcome variables. The results of these analyses are presented in Tables 22 through 29.

Note that the EIB subscales reported in previous analyses are not included. Data from non-IRA and experienced IRA teachers was not scored by subscales. The EIB-Part 1A score includes subscales I, III and 12 items in subscale IV. The EIB-Part 1B score includes 12 additional items from subscale IV and subscales V and VI.

TABLE 22: Comparison of IRA Blue and Yellow Version Teachers with Non-IRA Yellow Version Teachers on EIB-1A Posttest Student Mean Scores

GROUP	N	MEAN	S.D.	F RATIO	P
IRA - Yellow	607	20.71	4.89	20.41	.0000
Non-IRA - Yellow	307	18.33	6.08		
IRA - Blue	29	19.59	5.34		

TABLE 23: Comparison of IRA Blue and Yellow Version Teachers with Non-IRA Yellow Version Teachers on EIB-1B Posttest Student Mean Scores

GROUP	N	MEAN	S.D.	F RATIO	P
IRA - Yellow	592	41.35	7.73	22.38	.0000
Non-IRA - Yellow	294	37.33	9.95		
IRA - Blue	29	41.48	7.40		

TABLE 24: Comparison of IRA Blue and Yellow Version Teachers with Non-IRA Yellow Version Teachers on BSBI Subscale A (Curiosity) Posttest Student Mean Scores

GROUP	N	MEAN	S.D.	F RATIO	P
IRA - Yellow	435	2.73	.67	6.49	.0020
Non-IRA - Yellow	141	2.53	.73		
IRA - Blue	107	2.81	.66		

TABLE 25: Comparison of IRA Blue and Yellow Version Teachers with Non-IRA Yellow Version Teachers on BSBI Subscale B (Openness) Posttest Student Mean Scores

GROUP	N	MEAN	S.D.	F RATIO	P
IRA - Yellow	435	3.74	.67	18.59	.0000
Non-IRA - Yellow	141	3.37	.74		
IRA - Blue	107	3.79	.58		

TABLE 26: Comparison of IRA Blue and Yellow Version Teachers with Non-IRA Yellow Version Teachers on BSBI Subscale C (Satisfaction) Posttest Student Mean Scores

GROUP	N	MEAN	S.D.	F RATIO	P
IRA - Yellow	435	3.61	.68	5.18	.0061
Non-IRA - Yellow	141	3.46	.75		
IRA - Blue	107	3.74	.68		

TABLE 27: Comparison of IRA Blue and Yellow Version Teachers with Non-IRA Yellow Version Teachers on BSBI Subscale D (Responsibility) Posttest Student Mean Scores

GROUP	N	MEAN	S.D.	F RATIO	P
IRA - Yellow	435	3.90	1.05	9.07	.0003
Non-IRA - Yellow	141	3.58	1.03		
IRA - Blue	107	4.12	.94		

TABLE 28: Comparison of IRA Blue and Yellow Version Teachers with Non-IRA Yellow Version Teachers on BSBI Total Posttest Student Mean Scores

GROUP	N	MEAN	S.D.	F RATIO	P
IRA - Yellow	435	13.98	2.28	17.05	.0000
Non-IRA - Yellow	141	12.93	2.53		
IRA - Blue	107	14.46	1.92		

TABLE 29: Comparison of IRA Blue and Yellow Version Teachers with Non-IRA Yellow Version Teachers on CFE Posttest Student Mean Scores

GROUP	N	MEAN	S.D.	F RATIO	P
IRA - Yellow	558	21.22	6.25	39.63	.0000
Non-IRA - Yellow	310	24.17	6.39		
IRA - Blue	89	17.97	7.22		

Application of the Hartleys F_{max} test to each analysis demonstrated that the homogeneity of variance assumption underlying analysis of variance was satisfied in each case.

From Tables 22 to 29 it can be seen that all of the F ratios for comparing the three groups of teachers were significant beyond the .01 level of significance, indicating that there were significant differences among the posttest means for all of the student outcome variables. In order to determine which pairwise means were significantly different, the Newman-Keuls A Posteriori test was computed for all pairs of means. The results of this analysis are presented in Table 30.

TABLE 30: Table of Pairwise Differences at the .05 Level of Significance as Indicated by the Newman-Keuls A Posteriori Test

TEST	GROUP 1≠ / GROUP 2	GROUP 1 / GROUP 3	GROUP 2 / GROUP 3
EIB 1A	*		
EIB 1B	*		*
BSBI A	*		*
BSBI B	*		*
BSBI C			*
BSBI D	*		*
BSBI Total	*		*
CFE	*	*	*

≠ Groups: 1. IRA - Yellow Version
2. Non-IRA - Yellow Version
3. IRA - Blue Version

All of the comparisons of the IRA Yellow Version teachers' students with the non-IRA Yellow Version teachers' students were significant (P = .05) except for the BSBI sub scale C score. Of those comparisons showing a significant difference, the IRA Yellow Version teachers' students were significantly higher for all of these differences except for the CFE scores. On the CFE, the non-IRA Yellow Version teachers' students scored significantly higher than both the IRA Yellow Version and IRA Blue Version teachers' students, and the IRA Yellow Version students scored significantly higher than the IRA Blue Version students.

All of the comparisons of the students of the IRA Blue Version teachers with the non-IRA Yellow Version students were significant (P = .05) except for the EIB

1A scores. For those comparisons showing a significant difference, the students of the IRA Blue Version teachers were significantly higher than the non-IRA students in all comparisons except for the CFE scores. As noted above, the IRA Blue Version students were significantly below both the IRA and the non-IRA Yellow Version students on the CFE.

The only pairwise comparison between the IRA Yellow Version with the IRA Blue Version students that was significant was on the CFE. All of the other comparisons involving these two groups of students were not significant at the .05 level.

Interpretation: Despite the superiority by the control group in verbal and numerical ability as measured by the DAT, the IRA student groups had significantly superior posttest scores to the control group in cognitive inquiry and affective qualities of inquiry. These results were particularly meaningful for evaluating the effectiveness of the IRA program in light of the fact that the IRA program has been developed to operationalize the attitudinal and cognitive inquiry objectives delineated in Inquiry Objectives in the Teaching of Biology (Bingman et al., 1969). These results indicate that the IRA program is an effective teaching approach for developing cognitive inquiry skills and affective qualities of inquiry which have been previously recognized by science educators as important goals of science teaching.

Note that these results on the EIB and BSBI analyses also supported the validity of the IRA Yellow Version students' pre to post gains presented and discussed in Sub-Problem 1.

With respect to the posttest biology content instrument, CFE, student mean scores for the non-IRA - Yellow Version group significantly exceeded the scores for the IRA-Yellow Version Group. This finding should be interpreted in terms of the differences in the two student groups on DAT scores (Verbal and Numerical), the standard error of measurement reported in the CFE Manual, and the quantity of content coverage in the IRA Yellow Version Groups.

Part of this difference may be due to the significant differences in the DAT scores (Verbal and Numerical) reported on Tabel 22 which was significantly higher for the non-IRA Yellow Version group.

Another factor to consider is that the difference in the mean scores for the two groups (2.95) is within the standard error of measurement (3.1 to 3.3) reported in the CFE Manual.

It is possible that some of the difference in the obtained scores can be attributed to measurement error and does not represent "true" difference in the scores of the two groups.

Note, that the first two IRA themes treat 41 per cent of the chapters in the BSCS Yellow Version Text; the majority of IRA Yellow Version teachers completed only 11 per cent of Theme III activities. The low extent of biology content treatment indicated by this information, plus IRA teachers own statements that content treatment was reduced from previous years when IRA was not used, indicate that the lower CFE scores may be due in part to reduced biology content treatment. (Interviews of both IRA teachers and non-IRA teachers in previous IRA studies showed that IRA teachers treated at least 25 per cent fewer text chapters than non-IRA; it is reasonable to assume that this disparity of treatment also existed in the 1972-73 field test study.)

In light of the probable disparity of content treatment and differences in CFE posttest scores, it can be implied that in using the IRA program and in thereby expanding course objectives to include cognitive inquiry and affective qualities development, teachers must be aware that some reduction in the scope of biology content treated may be necessary. It should be pointed out, however, that in previous studies (1969-79, Bingman, et al., 1970, p. 30; 1971-72, unpublished data) IRA classes scored significantly higher on CFE posttests than non-equivalent non-IRA classes; groups with equivalent DAT scores were used in these studies.

The Yellow Version IRA groups scored significantly higher than Blue Version IRA groups on CFE scores. There appeared to be no particular reason to believe that differences in DAT scores, measuring error, or differences in the treatment of subject matter coverage in the course should account for these differences. Also previous experience in studies conducted in local IRA Blue Version classes have shown that the students scored much higher than found in this study.

Part of the difference can probably be attributed to fifty per cent of the students included in the Blue Version sample for CFE being 9th graders. Based on previous experience with 9th grade students the investigators as well as the CFE Manual authors have found considerable differences in scores favoring 10th graders. Otherwise, the differences in these results remain unexplained.

In summary, the students of IRA Yellow Version teachers have shown significantly higher posttest scores on instruments measuring cognitive inquiry skills and affective affective of inquiry than students of non-IRA Yellow Version teachers. This suggests that the IRA program is an effective teaching methodology for the development of cognitive inquiry and affective qualiteis of inquiry. Students of non-IRA Yellow Version teachers have shown significantly higher posttest scores on an instrument measuring biology content knowledge than students of IRA Yellow Version teachers. This difference may be due in part to non-equivalent verbal and numerical abilities of the IRA and non-IRA students, error in measurement and to the probable disparity in biology content treated in the IRA and non-IRA classes. This result is also not consistent with results of two previous studies.

The Yellow Version IRA classes have shown significantly higher posttest scores on the CFE instrument than the Blue Version IRA classes. Other than the grade level difference in the two groups, the results appear inconsistent with past studies.

SUB-PROBLEM 5: What revisions in the program materials are indicated by the results of testing, student feedback, and teacher feedback?

Feedback was received via the Teacher's Log (a report filed by each teacher after each activity), other written correspondence from teachers, telephone conversations, on-site visits by program staff with teachers and students, student feedback on Views and Preferences - C (Seymour and Bingman, 1973) and Class Activities Questionnaire (Steele, et al., 1971), and testing data. Information from all sources was summarized by IRA staff; all summaries were checked by two other staff members to insure that agreement was reached that the summary conveyed the major issues from the original communication. V&P-C, CAQ, and testing data were represented by the analyses which have been reported here and in the other papers of this paper set.

A copy of the Teacher's Log instrument (reduced one-half) is shown below. When teachers reported information via other means (other written communications, telephone conversations, etc.), the information provided was generally similar to that asked for on the Teacher's Log.

ACTIVITY NO. _____

1. ACTIVITY COMPLETED: Yes No

   If any part or all of the activity was used, circle yes. If no part of the activity was used, circle no.

2. IN-CLASS TIME SPENT ON ACTIVITY: ____ minutes

   Indicate time in minutes to the nearest ten minutes that you and your students spent in class on this activity.

3. MODIFICATIONS IN ACTIVITY PROCEDURES: Yes No

   If you followed the procedures without any modifications, circle no; if you modified any part or omitted a part, circle yes.

4. EXPLAIN THE MODIFICATIONS YOU MADE AND WHY:

5. GENERAL REACTIONS:

   Give any reactions you have to the activity, training or the program requirements. Include your opinions on the activity sequence--should it have been followed or preceded by another activity, would you suggest another sequence?

A. SPECIFIC REACTIONS TO PRE- AND IN- CLASS INSTRUCTIONS:

B. SPECIFIC REACTIONS TO STUDENT MATERIALS:

6. PERCENTAGE OF STUDENTS MEETING CRITERIA FOR OBJECTIVES:

   Estimate the percentage of students who reached the criteria specified in the objectives.

7. HOW COULD THIS ACTIVITY BE IMPROVED?

   Suggest how this activity could be improved to better meet the specified objectives or objectives you would include.

Interpretation: Of the 36 activities for which feedback was summarized, fourteen required major changes. (There were eight activities for which too little feedback was received to meaningfully evaluate; two additional activities were designed only for data collection purposes).

It should be noted, however, that changes generally dealt with better directions to the teacher (more direction to execute activity, more accurate time estimates, more complete discussion of expected student outcomes or assessment of outcomes, etc.) or changes in clarity or usefulness in student materials (shorten student forms, clarify statements, etc.). Recommendations to delete activities or major parts of activities, to redirect the activity to new goals, to substitute other activities, etc., were only suggested in response to the introductory activities, 101 to 105. Even when such changes were suggested, common elements of an initial orientation to the IRA program were found in all teacher suggested revisions. The specific changes recommended have been previously reported (Seymour, et al., 1973).

Aside from the specific activity-by-activity recommendation, two general guidelines for revision resulted from the teacher feedback and testing data: (1) As much as is possible, reduce the number of student forms. Specific suggestions for deletion were seldom given. However, teachers felt that the number of student forms was overwhelming to students as well as difficult for teachers to manage from a simple logistics viewpoint. (Student forms have been reduced from 213 pages to 110; note, however, that about 60% of this reduction is due to changes in format and printing.) (2) More complete treatment of the biology text content is desirable. Low scores (of IRA students compared to non-IRA students) in biology content appeared to be attributable, at least in part, to the fact that few IRA teachers completed the program. Thus several sections of the text were not directly treated in the instructional materials used. This too has been addressed in revision. IRA activities have been reduced from 46 to 41, with at least one activity being optional. A larger portion of text material is directly referred to in Themes I and II activities. In addition, the clarification and simplification of both student forms and teacher directions should enhance more rapid completion of program activities.

In general, feedback suggests that the IRA materials were found to be adequate for implementation in the classroom and satisfactory to teachers in terms of usability.

Summary:

Student outcomes in the 1972-73 Inquiry Role Approach field test are reported and discussed. Using the Comprehensive Final Examination, Exploration in Biology-Topic 1, and Biology Student Behavior Inventory, Inquiry Role Approach students were found to make significant pre-to-post gains in biology content understanding, cognitive inquiry skills, and in curiosity, openness, and responsibility. In addition, student mean scores on the Social Skills Checklists and Attitude Checklists met the criterion levels, indicating attainment of desired social skills and attitudes.

Student outcome differences for students grouped in quartiles based on verbal reasoning and numerical ability aptitudes (Differential Aptitude Test) generally showed differences between students in low and high quartiles for scores on the Exploration in Biology-Topic 1 (measuring cognitive inquiry skills) but not for scores on the Biology Student Behavior Inventory (measuring attitudes) or the Comprehensive Final Examination (measuring biology content understanding).

Student outcome differences between Inquiry Role Approach students and non-Inquiry Role Approach students indicated significantly greater development of cognitive inquiry skills and attitudes for IRA students over non-IRA, and significantly higher biology content understanding for non-IRA students over IRA.

Data from teachers, students, and testing results have been used to revise the IRA instructional materials. The data used and generally characterization of the revisions are discussed.

REFERENCES

Bennett, G. K., Seashore, H. G., and Wesman, A. G. Differential aptitude tests. New York: The Psychological Corporation, 1959.

Bennett, G. K., Seashore, H. G., and Wesman, A. G. Manual for the Differential Aptitude Tests. New York: The Psychological Corporation, 1966.

Bingman, R. M., Ed., Anderson, J. R., Blankenship, J. W., Carter, J. L., Cleaver, T. J., Jones, W. G., Kennedy, M. H., Klinckmann, E., Koutnik, P. G., Lee. A. E., and Stothart, J. R. Inquiry objectives in the teaching of biology. Kansas City, Missouri: Mid-continent Regional Educational Laboratory, 1969.

Bingman, R. M., Koutnik, P. G., Neff, F., Havlicek, L. L., Koran, J. J., Jr., McClelland, N., and Stothart, J. Learning through inquiry: a social design. Kansas City, Missouri: Mid-continent Regional Educational Laboratory, 1970.

Bingman, R. M., and Koutnik, P. G. A small-group study approach for biology based on inquiry objectives. American Biology Teacher, 1970, 32, 548.

Bingman, R. M., Koutnik, P. G., Seymour, L. A., Padberg, L. F., and Bingman, K. A., Inquiry Role Approach. Teachers Manuals. Kansas City, Missouri: Mid-continent Regional Educational Laboratory, 1972.

Biological Sciences Curriculum Study. Comprehensive final examination. New York: The Psychological Corporation, 1965.

Biological Sciences Curriculum Study. Biological science. An inquiry into life. New York: Harcourt, Brace and World, 1968a.

Biological Sciences Curriculum Study. Molecules to man. Boston: Houghton Mifflin Company, 1968b.

Campbell, D. T., and Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally & Co., 1963.

Koos, E. M., Burmester, M. A., Garth, R. E., and Stothart, J. R. Explorations in biology: Topic 1. Bird population. Kansas City, Missouri: Mid-continent Regional Educational Laboratory, 1972.

Koutnik, P. G. Developing inquiry skills: Focus on high school biology. AETS Newsletter, 1970, 4, 52.

Krejcie, R. V., and Morgan, D. W. Determining sample size for research activities. Educational and Psychological Measurement, 1970, 30, 607.

National Educational Association. Small-sample techniques. The NEA Research Bulletin, 1960, 38, 99.

Seymour, L. A., and Bingman, R. M. Development of Views and Preferences - C. A paper presented to the annual meeting of the National Association for Research in Science Teaching, Detroit, 1973.

Seymour, L. A., Bingman, R. M., Koutnik, P. G., Padberg, L. F., Havlicek, L. L., Kocher, A. T., and Burton, K. A. Inquiry Role Approach field test report (1972-73). Kansas City, Missouri: Mid-continent Regional Educational Laboratory, 1973.

Seymour, L. A., Padberg, L. F., Bingman, R. M., Koutnik, P. G., and Burton, K. A. The measurement of program implementation and students' cognitive, affective, and social performance in a field test of the Inquiry Role Approach (1972-73) I. Implementation: Its documentation and relationship to student inquiry development. A paper presented to the annual meeting of the National Association for Research in Science Teaching, Chicago, 1974a.

Seymour, L. A., Bingman, R. M., Koutnik, P. G., and Padberg, L. F. The measurement of program implementation and students' cognitive, affective, and social performance in a field test of the Inquiry Role Approach (1972-73). II. II. Measurement of social skills and attitudinal development of IRA biology students. A paper presented to the annual meeting of the National Association for Research in Science Teaching, Chicago, 1974b.

Steele, J., House, E., and Kerins, T. An instrument for assessing instructional climate through low inference student judgments. American Educational Research Journal, 1971, 8, 449.

Steiner, H. E. A study of the relationship between teacher practices and student performance of selected inquiry process behaviors in the affective domain in high school biology classes. Unpublished doctoral dissertation, University of Texas, Austin, 1970.